Tardiness Bounds for EDF Scheduling on Multi-Speed Multicore Platforms*

Hennadiy Leontyev and James H. Anderson
Department of Computer Science, University of North Carolina at Chapel Hill
$\{$*leontyev,anderson*$\}$*@cs.unc.edu*

Abstract

Multicore platforms, which include several processing cores on a single chip, are being widely touted as a solution to heat and energy problems that are impediments to single-core chip designs. To accommodate both parallelizable and inherently-sequential applications on the same platform, heterogeneous multicore designs with faster and slower cores have been proposed. In this paper, we consider the problem of scheduling soft real-time workloads on such a platform.

1. Introduction

Given the thermal and power problems that plague single-processor chip designs, most major chip manufacturers are investing in multicore technologies to achieve higher system performance. A number of systems with a modest number of cores are currently available, and in the coming years, the number of cores per chip is expected to increase significantly. In fact, chips with hundreds of cores are envisioned. It remains to be seen, however, whether such an extensive degree of parallelism can be effectively exploited. Indeed, many applications exist that are *inherently* sequential. In light of this, one approach, which is being advocated by many chip designers, is to provide a mix of faster and slower cores on the same platform [7]. On such a platform, an inherently-sequential application would benefit by executing on a faster core, while parallizable applications could execute across many slower cores. While having only fast cores would obviously be desirable, faster cores require greater chip area, so such an approach would adversely limit the number of cores per chip.

Large multicore platforms will likely be used in settings where timing constraints are required. For example, one envisioned use of such platforms is as a central server within the home that multiplexes interactive applications that require best-effort service with multimedia applications that have real-time requirements [3]. Such applications might include streaming movies over the Internet, providing cable television, or executing custom programs such as video games. Timing constraints in these applications are typically *soft*: missed deadlines, though undesirable, are usually not disastrous. Such constraints are far more common than hard constraints in many settings [8]. Unfortunately, prior work on scheduling real-time workloads on heterogeneous multiprocessors has focused only on hard real-time systems. While such work can be applied to schedule soft real-time applications, this comes at the price of overly conservative system designs.

In this paper, we show that such conservatism can be eliminated if deadline misses are permissible. We show this by presenting an algorithm for multi-speed systems called EDF-ms, which is a variant of the global EDF (GEDF) scheduling algorithm. Like GEDF [4], deadline tardiness under EDF-ms is bounded when scheduling sporadic tasks. Further, such bounds do not require severe caps on total utilization. In contrast, even with same-speed cores, if deadlines cannot be missed, then caps that can approach 50% of the available processing capacity are required under all known scheduling algorithms, except for Pfair algorithms [2]. Pfair algorithms, which have not been studied in the context of multi-speed systems, schedule tasks one quautum at a time, and thus preempt and migrate tasks frequently. EDF-ms preempts and migrates tasks less frequently, does not require same-speed cores, and can accommodate tasks with high execution costs for which utilization exceeds one on slower cores.

Prior work. Work on scheduling in heterogeneous multiprocessor real-time systems was initiated by Funk and colleagues, who presented a number of scheduling algorithms and associated analysis methods for systems with hard timing constraints. References concerning this work can be found in Funk's Ph.D. dissertation [6] (and are not included here due to space constraints). As noted earlier, our emphasis on *soft* real-time systems distinguishes our work from

*Work supported by a grant from Intel Corp., by NSF grants CNS 0408996, CCF 0541056, and CNS 0615197 and by ARO grant W911NF-06-1-0425.

these earlier efforts.

In work that is more experimental in nature, Kumar *et al.* [7] measured throughput and job response times on a two-speed multicore system with partitioned scheduling, and presented dynamic load-balancing heuristics that maximize throughput. This work is of relevance to research on soft real-time systems, as job response times are considered, but it does not include any analysis for validating deadlines or deadline tardiness.

Our algorithm, EDF-ms, has been devised by utilizing ideas from two prior papers concerning symmetric multiprocessor systems by Devi *et al.* [1, 5]. In [1], an EDF-based algorithm called EDF-fm is presented that limits task migrations without restrictive caps on total utilization. EDF-fm is a hybrid of partitioned EDF and GEDF. In EDF-fm, tasks are categorized as either "fixed" or "migrating" (hence the suffix "fm"). A *fixed* task exclusively executes on a specific processor. On the other hand, each *migrating* task executes on two processors, with each of its invocations executing exclusively on one of its assigned processors. Individual task utilizations must be capped at 1/2, but total utilization is not restricted (other than being at most the system's total processing capacity). Invocations of each migrating task are distributed between its assigned processors so that these processors are not overloaded in the long run; however, short-term overloads are possible. Such overloads can cause fixed tasks to miss their deadlines. However, such misses are by bounded amounts only.

Because EDF-ms is a variant of GEDF, tardiness bounds established for GEDF are of relevance to our work. Such bounds were first established by Devi and Anderson in [4]. These bounds apply to any sporadic task system with total utilization at most M scheduled on M symmetric processors. Any task can have maximum tardiness. Such behavior might not be acceptable for certain applications. In [5], the analysis in [4] is extended to allow up to M "privileged" tasks to have any predefined tardiness value, including zero. The resulting variant of GEDF is called EDF-hl. (The suffix "hl" signifies that privileged tasks are given higher priority, and others lower.)

Our approach. In the problem considered herein, cores are organized into groups, where cores in the same group have the same speed. In the earlier EDF-fm algorithm, migrating tasks are prioritized over fixed tasks to ensure that the former have zero tardiness. This allows schedulability to be analyzed on each processor independently. (If a migrating task were to miss a deadline on one of its processors, then this might delay its next invocation on its other processor. As a result, the two processors could not be analyzed independently.) We desire to maintain a similar independence property across *groups* of cores. To do this, we categorize tasks as either "fixed" or "intergroup." A *fixed* task executes only on the cores in one group, while an *intergroup* task may be executed on two groups of cores. We use the term "intergroup" instead of "migrating" because a fixed task in our case may migrate (among the cores in its group). We distribute the invocations of an intergroup task between its two assigned core groups in the same way as invocations of migrating tasks are distributed in EDF-fm. Further, we treat intergroup tasks specially when scheduling, as in the earlier EDF-hl algorithm, so that they can be guaranteed zero tardiness. This enables each group of cores to be analyzed independently. However, one key difference arises in our analysis: the distribution pattern used for intergroup tasks allows short-term overloads to occur (for the same reason that such overloads occur in EDF-fm). Thus, the analysis of tardiness in [5] must be adjusted to allow privileged tasks to create short-term overloads.

Summary of contributions. The main contributions of this paper include devising EDF-ms and establishing tardiness bounds for it. In addition, we present an experimental evaluation of EDF-ms's effectiveness in limiting tardiness. To our knowledge, EDF-ms is the first algorithm proposed for multi-speed platforms that can schedule soft real-time tasks with bounded tardiness without severe utilization restrictions. Because the ideas underlying EDF-ms were originally proposed in completely different settings, new analysis for integrating these ideas had to be devised. EDF-ms, its analysis, and evaluation are presented in detail in Secs. 3–5, after first presenting our system model in Sec. 2.

2. System Model

We consider the problem of scheduling a set of sporadic tasks on $M \geq 4$ cores of $g \geq 2$ speeds. We will group cores by speed: we let m_h denote the number of cores in Group h, where $1 \leq h \leq g$, and we let s_h denote their speed. We assume $s_1 = 1$ and $s_j < s_k$ if $j < k$. We also assume that $m_h \geq 2$ holds for each h. (Given that our main focus is large multicore platforms, this is a reasonable assumption. However, we briefly consider later how to handle groups with only one core.)

We let τ denote the sporadic task system to be scheduled, and assume that it consists of n independent tasks, $T_1, \ldots, T_n$. Each task is invoked or *released* repeatedly, with each such invocation called a *job*. Associated with each task T_i are two parameters, e_i and p_i: e_i gives the maximum *execution time* of one job of T_i on a unit-speed core, while, p_i, called the *period* of T_i, gives the minimum time between two consecutive job releases of T_i. On a core with speed s_h, a job of T_i completes in e_i/s_h time units. For brevity, T_i's parameters are sometimes denoted using the notation $T_i = (e_i, p_i)$.

The k^{th} job of T_i, where $k \geq 1$, is denoted $T_{i,k}$. A task's first job may be released at any time at or after time zero.

The release time of the job $T_{i,k}$ is denoted $r_{i,k}$ and its (absolute) deadline $d_{i,k}$ is defined as $r_{i,k} + p_i$. Each task is sequential, so at any time, it may execute on at most one core. When a job of a task misses its deadline, the release time of the next job of that task is unaltered. This ensures that each task receives a processor share in accordance with its utilization (defined below) in the long term. Thus, a task may release a new job when prior job(s) of that task have not been completed. Such a new job cannot commence execution until the prior jobs have completed. If a job $T_{i,j}$ with a deadline at $d_{i,j}$ completes at time t, then its *tardiness* is defined as $max(0, t - d_{i,j})$. A *task*'s tardiness is the maximum of the tardiness of any of its jobs.

The *utilization of task* T_i is defined as $u_i = e_i/p_i$, and the *utilization of the task system* τ as $U_{sum} = \sum_{T_i \in \tau} u_i$. We require $\sum_{u_i > s_j} u_i \leq \sum_{k>j} m_k \cdot s_k$ and $U_{sum} \leq \sum_{h=1}^{g} m_h \cdot s_h$. Otherwise, tardiness can grow unboundedly. Note that the first of these requirements implies $u_i \leq max(s_j)$. Note also that it is possible that $u_i > s_j$ holds for some j.

In this paper, we assume that time is continuous, but execution costs, periods, and core speeds are rational.

3. Algorithm EDF-ms

The name EDF-ms stands for EDF *multi-speed*. Like EDF-fm [1], mentioned earlier, the algorithm consists of two phases: an offline *task assignment* phase and an online *execution* phase. When the task assignment phase is applied to some task set τ, at most g groups of tasks are created. In addition, there may be up to $g - 1$ tasks that do not belong to any group. Each of these tasks may execute on the cores of two groups. For each pair of consecutive groups, say Group h and Group $h + 1$, at most one task that migrates between them may exist. We denote this task (if it exists) as $T^{h,h+1}$ and call it an *intergroup* task. All other tasks are called *fixed* tasks, as each executes within one group only. The assignment algorithm (which is not shown, due to space constraints) sorts tasks by utilization and assigns tasks to groups by exhausting the capacity of faster groups first.

The assignment algorithm returns a set of values $Z_{k,h}$, where $1 \leq k \leq n$ and $1 \leq h \leq g$. The value $Z_{k,h}$ denotes the fraction of T_k's utilization that is assigned to Group h. For any fixed task T_k assigned to Group h, $Z_{k,h} = u_k$ holds. For any intergroup task $T_k = T^{h,h+1}$, $Z_{k,h} + Z_{k,h+1} = u_k$ holds. If task T_k may not execute on Group h, then $Z_{k,h} = 0$. The assignment algorithm ensures the following.

$$\sum_h Z_{i,h} = u_i \quad \wedge \quad \sum_i Z_{i,h} \leq m_h \cdot s_h \tag{1}$$

Example 1. Fig. 1 shows the assignment determined for the task set $\tau = \{T_1, \ldots, T_{13}\}$. For conciseness, we will use the notation T_{i-j} to denote $T_i, \ldots, T_j$ in describing this task set. τ consists of $T_{1-5} = (8, 10)$, $T_{6-9} = (3, 2)$, and $T_{10-13} = (4, 2)$. Tasks T_{1-5}, T_{6-9}, and T_{10-13} have utilization 0.8, 1.5, and 2.0, respectively. The system is comprised of three groups of cores of speeds one, two, and three. That is, $m_1 = 3$, $s_1 = 1$, $m_2 = 3$, $s_2 = 2$, $m_3 = 3$, and $s_3 = 3$. In Fig. 1, each core group is depicted as a "bin" with its height proportional to $m_h \cdot s_h$, *i.e.*, the total processing capacity of the group. As seen, tasks T_{1-3} are assigned to Group 1, tasks T_{5-8} to Group 2, and tasks T_{10-13} to Group 3. Tasks $T_4 = T^{1,2}$ and $T_9 = T^{2,3}$ are intergroup tasks.

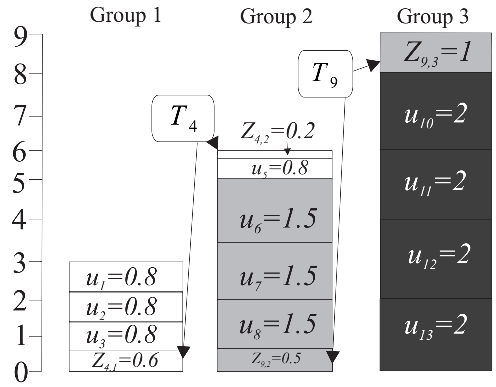


Figure 1. Mapping of the task set onto groups of cores in Example 1. Different shadings are used for tasks with utilization at most one, two, and three.

For each Group h, we can define a set of tasks τ_h with jobs to be scheduled by this group:

$$\tau_h = \{T_i : Z_{i,h} > 0\}.$$

As with migrating tasks in EDF-fm [1], if an intergroup task $T^{h,h+1}$ were to miss its deadline in the schedule for either Group h or Group $h + 1$, then this would create a nontrivial linkage between these two groups that complicates scheduling analysis. This is because, if a job of $T^{h,h+1}$ misses its deadline, then the processing of the next job of $T^{h,h+1}$ may be delayed until *after* its release, and this may increase the chance that it will miss as well. Thus, missed deadlines in one group could lead to missed deadlines in another group. Thus, our scheduling policy must achieve two goals: **(i)** allow us to analyze the schedule in each group independently, and **(ii)** not overload any group in the long run (for otherwise, tardiness in such a group would grow unboundedly).

These goals are accomplished as follows. During the execution phase of EDF-ms, jobs of tasks in τ_h are scheduled

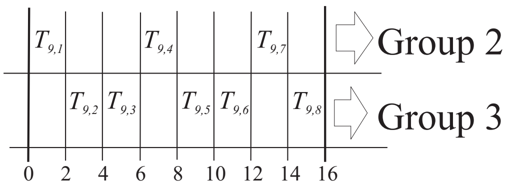


Figure 2. Assignment of jobs of the task T_9 from Example 2.

on the cores in Group h using GEDF, with the jobs of intergroup tasks treated specially. As we shall see, the special treatment given to intergroup tasks ensures that their jobs always have zero tardiness. This allows us to analyze each core group as a separate (same-speed) system. The jobs of each intergroup task $T_k = T^{h,h+1}$ are distributed between its assigned Groups h and $h+1$ using a special deterministic pattern first described in [1], which ensures that the total workload from these jobs assigned to these two groups over the long term is in accordance with the shares $Z_{k,h}$ and $Z_{k,h+1}$, respectively.

In order to describe this assignment pattern, we introduce some additional notation. For each intergroup task $T_k = T^{h,h+1}$, we let $f_{k,h}$ and $f_{k,h+1}$ denote the fraction of T_k's jobs that are processed by Groups h and $h+1$, respectively. These two quantities are defined as follows.

$$f_{k,h} = Z_{k,h}/u_k \quad \wedge \quad f_{k,h+1} = Z_{k,h+1}/u_k$$

Note that, by (1), $f_{k,h} + f_{k,h+1} = 1$.

To explain the assignment pattern, we consider a single group, Group h. Assume that there exist two intergroup tasks $T^{h-1,h}$ and $T^{h,h+1}$. If we were to depict Group h as a "bin" as done in Fig. 1, then $T^{h-1,h}$ would be the *top* task in Group h, and $T^{h,h+1}$ would be the *bottom* task. Different (complementary) assignment rules are needed for these two cases. Let T_k denote either $T^{h-1,h}$ or $T^{h,h+1}$. Let $j \geq 1$ be the index of the latest job released by T_k at or after time t and let j_a be the number of jobs of T_k assigned to Group h before t. For the case where $T_k = T^{h-1,h}$, *i.e.*, T_k is the top task, the j^{th} job of T_k is assigned to Group h iff $j - 1 \neq \lfloor j - j_a/(1 - f_{k,h}) \rfloor$. We call this assignment rule the *top rule*. For the case where $T_k = T^{h,h+1}$, *i.e.*, T_k is the bottom task, the j^{th} job of T_k is assigned to Group h iff $j - 1 = \lfloor j_a/f_{k,h} \rfloor$. We call this assignment rule the *bottom rule*.

Example 2. Fig. 2 shows the assignment pattern for the jobs of task T_9 from Example 1, which has frequencies $f_{9,2} = Z_{9,2}/u_9 = 0.5/1.5 = 1/3$ and $f_{9,3} = Z_{9,3}/u_9 = 1/1.5 = 2/3$. Jobs of T_9 are assigned to Group 2 using the bottom rule and to Group 3 using the top rule. Here, we will focus on Group 2. Consider the time instant $t = 6$ when the fourth job of T_9 is released. When considering this fourth job in the bottom rule, $j = 4$. Prior to time t, one job of T_9 was assigned to Group 2, so $j_a = 1$. Applying these values via the bottom rule, we obtain $j - 1 = 3 = \lfloor 1 \cdot 3 \rfloor = \lfloor j_a/f_{9,2} \rfloor$, so the job is assigned to Group 2. Note that, by time 6, three jobs of T_9 are released and one of them is processed by Group 2, which is in accordance with the fraction $f_{9,2} = 1/3$.

As shown in [1], this assignment strategy ensures that the maximum number of jobs of an intergroup task T_k released during an interval of length t and assigned to Group h is at most $\lceil f_{k,h} \lceil \frac{t}{p_k} \rceil \rceil$. Thus, the maximum demand due to jobs of T_k that must be processed by Group h during an interval of length t is at most $\left\lceil f_{k,h} \left\lceil \frac{t}{p_k} \right\rceil \right\rceil e_k$, which is approximately $Z_{k,h} \cdot t$. Because the demand of each fixed task $T_k \in \tau_h$ during the interval $[0, t)$ is at most $u_k \cdot t$, and the demand of each intergroup task T_k assigned to Group h is approximately $Z_{k,h} \cdot t$, Group h will not be overloaded in the long run. For example, for Group 2 in Fig. 1, these values sum to $6t$, which matches the group's overall computing capacity within $[0, t)$, as given by $m_2 \cdot s_2 \cdot t$.

Because no group is overloaded in the long term, the scheduling policy we give below for each group will ensure that the jobs of intergroup tasks never miss their deadlines. As such, we no longer need to consider multiple groups, but can concentrate our analysis efforts on just one, say Group h. Furthermore, Group h's per-core speed of s_h is no longer an issue, since all cores in the group have the same speed. We therefore assume that all cores in Group h have a speed of one and that all execution costs, utilizations, and $Z_{k,h}$ values of tasks executing on the cores of Group h have been scaled by dividing them by s_h.

We further simplify the problem notationally by assuming that we have have m (unit-speed) processors upon which we must schedule a set of $n+2$ sporadic tasks, $\tau = \{T_0, T_1, \ldots, T_{n+1}\}$. T_0 and T_{n+1} represent, respectively, the top and bottom intergroup tasks for this core group. (Later, we explain how to adjust our results if either of these tasks does not exist.) $T_1, \ldots, T_n$ are the fixed tasks for the group. Our scheduling policy treats T_0 and T_{n+1} specially so that their jobs do not miss their deadlines, so we call them *privileged* tasks. Jobs of the privileged tasks are assigned to the system using the top and bottom rules discussed earlier. We let $Z_0 < u_0$ and $Z_{n+1} < u_{n+1}$ denote the part of the utilization of tasks T_0 and T_{n+1}, respectively, that must be processed by the system. If T_0 and T_{n+1} both exist, as assumed here, then $Z_0 + Z_{n+1} + \sum_{i=1}^{n} u_i = m$. More generally, $Z_0 + Z_{n+1} + \sum_{i=1}^{n} u_i \leq m$. We let $f_0 = Z_0/u_0$ and $f_{n+1} = Z_{n+1}/u_{n+1}$.

Jobs of privileged tasks are treated specially in scheduling by using an approach presented in [5]. In this approach, the concept of slack is used: if job $T_{k,j}$ executes for $\delta_{k,j}$ time prior to time $t \leq d_{k,j}$, then its *slack* at t, given by $d_{k,j} - t - (e_k - \delta_{k,j})$, represents the maximum amount of time that $T_{k,j}$ can remain idle (*i.e.*, not execute) and still meet its deadline. This concept is used in scheduling tasks in the following way: all jobs are scheduled using GEDF, with the exception that, if a job of a privileged task has zero slack, then it is executed continuously until its deadline. This policy clearly ensures that privileged tasks do not miss their deadlines. (Recall that each core group consists of at least two cores.)

4. Tardiness Bounds

We describe the tardiness-bound derivation for EDF-ms for tasks in one group, which is based upon the techniques from [4, 5]. Due to space constraints it is not possible to present it here in detail, these details can be found online at http://www.cs.unc.edu/~anderson/papers.

We say that the system is *non-concrete* if the release times of jobs are not specified, and *concrete*, otherwise. Given an arbitrary non-concrete task system τ^N, we want to determine the maximum tardiness of any job of any non-privileged task in any concrete instantiation of τ^N. Let $\tau_H = \{T_0, T_{n+1}\}$ be the set of privileged tasks, and let $\tau_L = \{T_1, \ldots, T_n\}$ denote the remaining tasks. Let τ be a concrete instantiation of τ^N. All jobs of non-privileged tasks are processed by the system, and the jobs of privileged tasks are processed according to top and bottom rules. Let $T_{\ell,j}$ be a job of a non-privileged task in τ, let $t_d = d_{\ell,j}$, and let $\mathcal{S}$ be an EDF-ms schedule for τ with the following property.

> The tardiness of every job of every non-privileged task T_k in τ with deadline less than t_d is at most $x + e_k$ in $\mathcal{S}$, where $x \geq 0$.

Our goal is to determine the smallest x, independent of the parameters of T_ℓ, such that the tardiness of $T_{\ell,j}$ remains at most $x + e_\ell$. Such a result would by induction imply a tardiness of at most $x + e_k$ for all jobs of every non-privileged task $T_k \in \tau$. Because τ is arbitrary, the tardiness bound will hold for every concrete instantiation of τ^N.

Assume that $T_{\ell,j}$ misses its deadline (for otherwise, its tardiness is zero). The completion time of $T_{\ell,j}$ then depends on the amount of work that can compete with $T_{\ell,j}$ after t_d. Hence, a value for x can be determined via the following steps.

- Compute an upper bound (UB) on the work pending for tasks in τ (including that due to $T_{\ell,j}$) that can compete with $T_{\ell,j}$ after t_d.
- Determine a lower bound (LB) on the amount of such work required for the tardiness of $T_{\ell,j}$ to exceed $x + e_\ell$.
- Determine the smallest x such that the tardiness of $T_{\ell,j}$ is at most $x + e_\ell$ using UB and LB.

The value of x so obtained is characterized in the following theorem.

Theorem 1. *Let $U(\tau, y)$ denote the set of at most y tasks of highest utilization from the task set τ. Let $E(\tau, y)$ denote the set of at most y tasks with the highest execution costs from τ. Let $E_L = \sum_{u_i \in E(\tau_L, m-1)} e_i$ and $U_L = \sum_{e_i \in U(\tau_L, min(m-2, |\tau_L|))} u_i$. Let $e_{max} = max(e_i)$ and $e_{min} = min(e_i)$, let*

$$x_1 = \frac{E_L + e_0 + \sum_{k \in \{0, n+1\}} e_k(1 + f_k - 2Z_k) - e_{min}}{m - |\tau_H| - U_L}, \quad (2)$$

$$x_2 = \frac{e_0 + E_L + \sum_{k \in \{0, n+1\}} e_k(3 - Z_k) + A}{m - max(|\tau_H| - 1, 0) max_{T_i \in \tau_L} u_i - U_L - Z_0 - Z_{n+1}}, \quad (3)$$

where

$$A = \begin{cases} (Z_0 + Z_{n+1} - 1)e_{min} & \text{if } Z_0 + Z_{n+1} \leq 1 \\ (Z_0 + Z_{n+1} - 1)e_{max} & \text{otherwise.} \end{cases}$$

With $x = min(x_1, x_2)$, tardiness for a non-privileged task T_k scheduled under EDF-ms is at most $x + e_k$.

Note that, for tardiness to be bounded under EDF-ms, the denominators in the right-hand-side expressions in (2) and (3) must not be zero. This gives us two requirements, $m - |\tau_H| - U_L > 0$, and $m - max(|\tau_H| - 1, 0)(max_{T_i \in \tau_L} u_i) - Z_0 - Z_{n+1} - U_L > 0$. Thus, to ensure bounded tardiness, some slight restrictions on task utilizations are required. (The impact of these restrictions is assessed in the next section.)

We have assumed above that there are two privileged tasks. If only one such task exists, then we can assume there are two, with the execution cost, utilization, and Z and f values for one of them being zero. Of course, if there are no privileged tasks, then tardiness can be analyzed using the results from [4].

We have also required that there be at least two cores per group. A group with one core can be handled in three ways.

- We can use the same approach described above, but limit the group to have at most one privileged task. In this case, the group's lone core may not be fully utilized.
- We can schedule the tasks within the group like EDF-fm schedules tasks on one processor. Note that EDF-fm requires that if two privileged tasks exist, then they have a combined share on the processor of at most one. On the other hand, the processor can be fully utilized.

- We can combine the lone core in the group with slower cores to create a group of at least two cores. This comes at the expense of not utilizing the full processing capacity of the core added to the slower group.

The best approach will depend on the workload to be scheduled.

5. Experimental Evaluation

In this section, we present an experimental evaluation of EDF-ms. We performed two sets of experiments. In the first, we assessed tardiness within a single group. In the second, we assessed the impact of several variants of the task-assignment method discussed in Sec. 2 on overall tardiness.

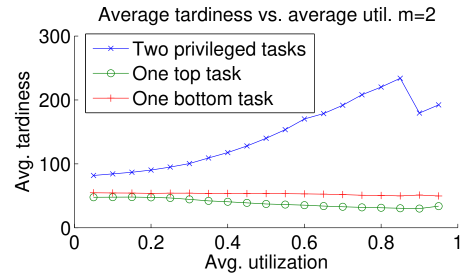
5.1. Tardiness Bounds for a Single Group

In this set of experiments, we computed per-task tardiness bounds for random task sets on $m = 2$, 4, 8, and 16 unit-speed processors in the presence of one or two privileged tasks (top or bottom). Each task set consisted of at least $m+1$ tasks. Tasks within each set were generated with utilizations uniformly distributed in $[0, u_{max})$, where u_{max} ranged from 0.1 to 1 in steps of 0.05. For each value of u_{max}, 1,000 task sets were generated. Task execution costs were uniformly distributed over $[10, 20)$. Tasks were added to the generated task set until total utilization exceeded m. The shares of the privileged tasks were then defined so that $Z_0 + Z_{n+1} + \sum_{i=1,\ldots,n} u_i = m$. The top (bottom) task (if either existed) was taken to be the task with the smallest (largest) utilization.

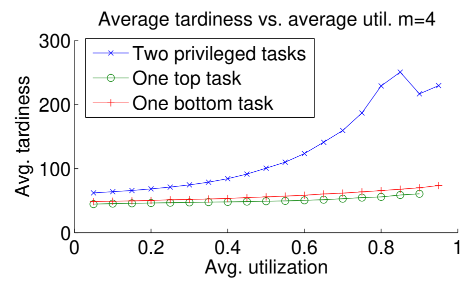
Fig. 3 shows the average *maximum* task tardiness plotted against the average task utilization, u_{avg}, for different values of m. Note that tardiness grows as u_{avg} grows, with the exception of the case of two processors and one privileged task, shown in inset (a). In this case, if we apply Theorem 1 for $m = 2$, then $U_L = 0$, and only one of Z_0 and Z_{n+1} is non-zero and it is at most one. In this case, the denominators of (2) and (3) are independent of the utilizations of non-privileged tasks.

The situation for $m = 2$ changes drastically if there are two privileged tasks. During some time intervals, both processors in the group must execute jobs of privileged tasks. During such intervals, non-privileged tasks cannot execute at all. If the number of non-privileged tasks is small, and they have high utilizations, then these tasks recover slowly from this shortage of processing capacity, as demand due to privileged tasks lessens. This situation is depicted in the right part of inset (a), where non-privileged tasks have high utilizations.

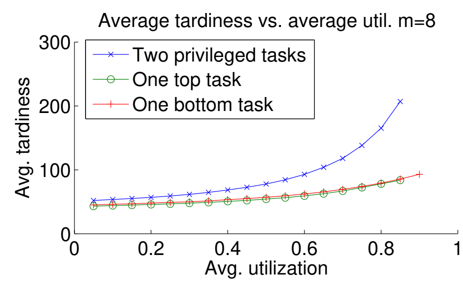
As seen in insets (b)–(d), this effect eases as the number of processors grows. This is because, with more than



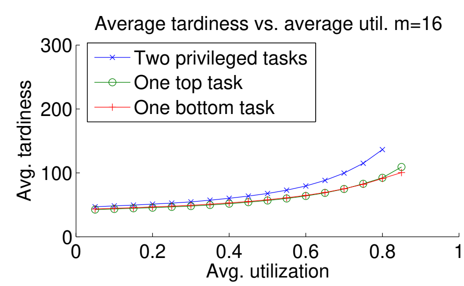
(a)



(b)



(c)



(d)

Figure 3. Tardiness bounds versus u_{avg} for (a) $m = 2$, (b) $m = 4$, (c) $m = 8$, and (d) $m = 16$.

two processors, at least one processor is always available to execute non-privileged tasks. As the number of processors grows, more processing capacity is available for executing the jobs of non-privileged tasks. Hence, tardiness decreases. This suggests that EDF-ms may be very effective in large multicore systems, the main focus of our work.

Note that the curves in insets (a) and (b) for the case of two privileged tasks do not continue to increase at the right. This is because, when m is only two or four but two privileged tasks exist, the number of samples with high u_{avg} is small.

As remarked earlier, bounded tardiness is guaranteed under EDF-ms only if the two conditions $m - max(|\tau_H| - 1, 0)max_{T_i \in \tau_L}(u_i) - Z_0 - Z_{n+1} - U_L > 0$ and $m - |\tau_H| - U_L > 0$ hold. These conditions are not very restrictive. As evidence of this, no task set generated in this set of experiments had to be rejected because of these conditions.

5.2. Task Assignments, Revisited

Because tardiness within a group depends on the parameters of the privileged tasks in that group, it might be possible to lower overall tardiness by using a task-assignment policy that lessens the impact of privileged tasks on other tasks. To see if this is so, we considered two such policies and compared them to the one described in Sec. 2. In that which follows, we refer to original policy described earlier as SIMPLE.

According to Theorem 1, if the privileged tasks within a group either require large shares within the group or have high execution costs, then tardiness within the group may be high. This suggests two alternative assignment policies, one that seeks to minimize the shares of privileged tasks, and a second that seeks to minimize their execution costs. Both policies function in a similar way: after running SIMPLE, consider the groups in decreasing index order, and for each group, select as the bottom task the task in the group with nonzero share that has the smallest utilization or execution cost. We call the former scheme MIN-UTIL, and the latter, MIN-EXEC. (If a task with a lower utilization is selected as the bottom task in Group h, then it may actually fit within Group $h + 1$. Thus, in both schemes, the process of assigning tasks to groups is in fact iterative.)

To evaluate the impact of MIN-UTIL and MIN-EXEC, we considered three system configurations C_1, C_2, and C_3, which have a small, medium, and large total number of cores M. Each configuration consists of cores with speeds one, two, and three. The number of cores of each type is shown in Table 1. For each configuration, we evaluated 60 task sets. The tasks in each set were generated as follows. First, tasks with utilizations distributed randomly in $[0, 2.1)$ were generated until the processing capacity of Group 3 would be exceeded. Then, tasks with utilizations in $[0, 1.4)$ were generated until the combined processing capacities of Groups 2 and 3 would be exceeded. Finally, the remaining tasks were generated with utilizations in $[0, 0.7)$ until the remaining capacity of the system was exhausted. All task execution costs were distributed uniformly over $[1, 100)$. For each generated task system, we used Theorem 1 to compute the maximum tardiness of the non-privileged tasks in each group under each assignment scheme.

	Core groups			
	$s_1 = 1$	$s_2 = 2$	$s_3 = 3$	M
C_1	12	4	2	18
C_2	24	8	4	36
C_3	48	16	8	72

Table 1. Three evaluated configurations.

Fig. 4 shows the maximum tardiness per group for each configuration. Each point in each group gives the maximum tardiness of one of the generated task sets. As the graphs show, MIN-EXEC results in significantly lower tardiness for Group 2 and slightly lower tardiness for the other groups. On the other hand, the use of MIN-UTIL did not result in better tardiness than SIMPLE. However, this could be an artifact of our task-generation methodology. The overall conclusion to be drawn from these results is that the significant flexibility that exists in the task-assignment process can be exploited to realize certain benefits in some systems. (In particular, this assignment process is not rigid like the bin-packing strategies used in partitioning schemes.) Other benefits beyond lowering tardiness are possible. For example, some hard real-time tasks could be supported by choosing them as intergroup tasks. Also, the response times of certain tasks could be lowered by assigning them to faster cores, as long as the resulting assignment is valid and utilization constraints are met.

6 Conclusion

We have presented a new algorithm, EDF-ms, which can be used for scheduling sporadic soft real-time task systems on asymmetric multicore platforms with cores of different speeds. To our knowledge, this paper is the first to propose a scheduling approach for such heterogeneous platforms that is suitable for soft real-time workloads that require bounded deadline tardiness. Our algorithm is capable of fully utilizing the processing capacity of the system, provided certain very slight restrictions on task utilizations hold. This property comes at the price of needing to migrate tasks, as required in global scheduling approaches such as GEDF. Note that the main cost of a migration is a loss of cache affinity. Thus, in a multicore platform, the need to migrate

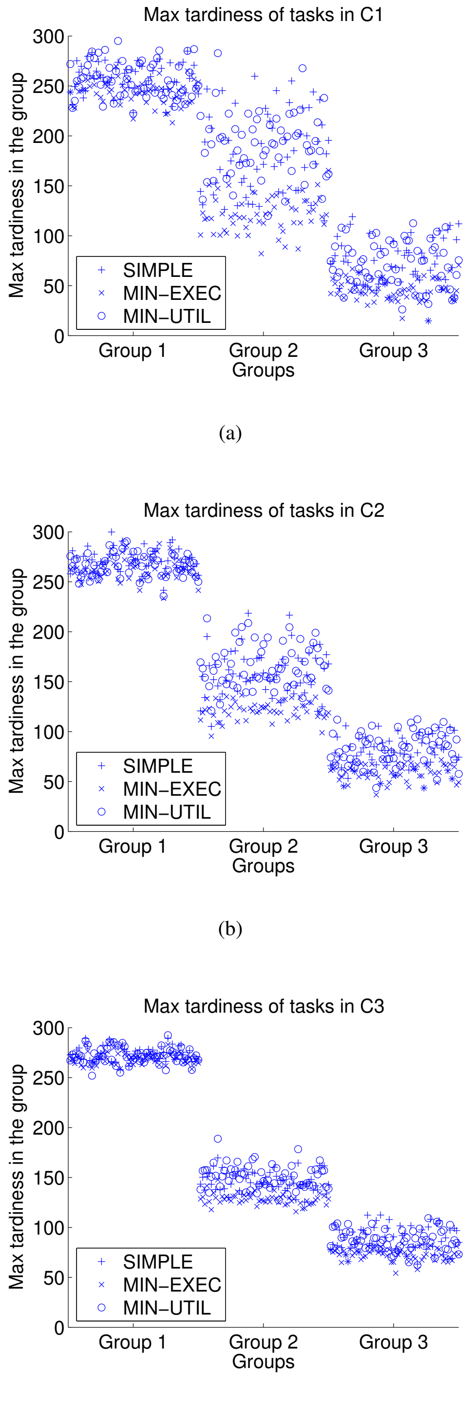


Figure 4. Tardiness bounds under different assignment schemes for three core groups in three different configurations: (a) C_1, (b) C_2, and (c) C_3.

tasks is less of a concern than for a traditional SMP, due to the presence of shared on-chip caches. Although we have not directly included migration costs in our task model, they can be accounted for by inflating task execution costs to include the cost of migrations, as is commonly done in real-time scheduling analysis.

Several interesting avenues for further work exist. For example, it would be interesting to extend our results to include tasks with synchronization requirements. It would also be interesting to consider workloads with both soft real-time and non-real-time tasks. Finally, in this paper we have considered heterogeneous platforms where the cores only differ in speed. This is different from *functional asymmetry*, where each core has a different set of "capabilities" and tasks must be matched with cores possessing the capabilities they need. It would be interesting to extend our results to apply to such platforms.

References

[1] J. Anderson, V. Bud, and U. Devi. An EDF-based scheduling algorithm for multiprocessor soft real-time systems. In *Proc. of the 17th Euromicro Conf. on Real-Time Systems*, pp. 199–208, 2005.

[2] S. Baruah, N. Cohen, C.G. Plaxton, and D. Varvel. Proportionate progress: A notion of fairness in resource allocation. *Algorithmica*, 15:600–625, 1996.

[3] J. Calandrino, D. Baumberger, T. Li, S. Hahn, and J. Anderson. Soft real-time scheduling on performance asymmetric multicore platforms. In *Proc. of the 13th IEEE Real-Time and Embedded Technology and Applications Symp.*, pp. 101–110, 2007.

[4] U. Devi and J. Anderson. Tardiness bounds for global EDF scheduling on a multiprocessor. In *Proc. of the 26th IEEE Real-Time Systems Symp.*, pp. 330–341, 2005.

[5] U. Devi and J. Anderson. Flexible tardiness bounds for sporadic real-time task systems on multiprocessors. In *Proc. of the 20th IEEE International Parallel and Distributed Processing Symp.*, 2006.

[6] S. Funk. *Implementing Real-time Systems on Heterogeneous Multiprocessors*. Ph.D. dissertation, The University of North Carolina at Chapel Hill, 2004.

[7] R. Kumar, D. Tullsen, P. Ranganathan, N. Jouppi, and K. Farkas. Single-ISA heterogeneous multi-core architectures for multithreaded workload performance. *Proc. of the 31st International Symp. on Computer Architecture (ISCA)*,pp. 64–75, 2004.

[8] R. Rajkumar. Resource Kernels: Why Resource Reservation should be the Preferred Paradigm of Construction of Embedded Real-Time Systems. Keynote talk, 18th Euromicro Conference on Real-Time Systems, Dresden, Germany, 2006.